Computers Math. Applic. Vol. 21, No. 8, pp. 1–7, 1991
Printed in Great Britain.
0097-4943/91 $3.00 + 0.00

AN APPROXIMATE SOLUTION TO THE PERT PROBLEM

D. SCULLI AND Y. W. SHUM
University of Hong Kong, Department of Industrial Engineering
Pokfulam Road, Hong Kong

(Received February 1990)

Abstract—The completion times of a set of paths of a PERT network are expressed as a multivariate normal distribution. Approximations are then proposed for computing the mean and standard deviation for the maximum from a set of multinormal variables. Results show the estimates to be very close to those obtained via simulation.

INTRODUCTION

The problem of finding the mean completion time of a PERT network when the distributions of individual activities are known is essentially one of statistical analysis. Difficulties arise because the durations of paths in a network are partially correlated, depending on the lengths of the common sections, and also because of the need to find the maximum of a set of variable duration times.

The research literature can be classified into four main groups. The first group is the research to find close lower and upper bounds to the completion time; examples include [1-3]. The second group consists of approximations in which the distributions of individual activities are assumed to be of the discrete type. These approximations involve the manipulation of a fixed number of time values and the corresponding probabilities; frequently quoted in this group are [4,5]. The third group also consists of approximations, where the computations involve the manipulation of only distribution parameters, and this research has usually assumed a normal distribution for individual activities; see [6-8]. The fourth group uses simulation to estimate the mean completion time and its distribution; examples include [9-11].

In this paper the completion time of a PERT network is formulated as the maximum from a set of variables that are distributed as a multivariate normal. The results of [6] on the greatest of a set of random variables are used to obtain an approximation to the mean and standard deviation of the completion time of the network.

FORMULATION OF COMPLETION TIME AS A MULTIVARIATE NORMAL

It is assumed that the durations of the individual activities of the network are independently distributed with known means and known variances. Consider n paths through the entire network, and let their duration times be given by the random variables $P_1, \dots, P_n$, respectively. The mean and variance, μ_i and σ_{ii}^2, of path P_i can be readily obtained by adding the respective means and variances of the activities that make up path P_i.

It is also assumed that each path P_i is normally distributed with mean μ_i and variance σ_{ii}^2. The normality assumption for the length of a path in a network is partly supported by the Central Limit Theorem, since it involves the addition of a number of independent activity duration times. There may be paths with only a few activities, in which case the normality assumption may not be totally justified, but these paths will be short in the majority of cases and may not significantly influence the average completion time of the network.

The covariance, σ_{ij}^2, of paths P_i and P_j can also be readily determined; see [6,12]. Paths P_i and P_j will have a set of activities that are common to both paths. If there are no activities in common, then the covariance will be zero, $\sigma_{ij}^2 = 0$. Let r_{ij} be the random variable representing the total duration of activities that are common to P_i and P_j, and let s_i and s_j be the random variables representing the total duration of the others or uncommon activities of paths P_i and P_j respectively, such that:

$$P_i = r_{ij} + s_i \quad \text{and} \quad P_j = r_{ij} + s_j .$$

Note that s_i, s_j and r_{ij} will not, by definition, have any common activities, and will therefore be independent. The covariance, σ_{ij}^2, will be given by

$$\sigma_{ij}^2 = \operatorname{Cov}\{P_i, P_j\} = \operatorname{Cov}\{r_{ij} + s_i, r_{ij} + s_j\} = \operatorname{Var}\{r_{ij}\}.$$

So, in effect, the covariance of paths P_i and P_j is simply equal to the sum of the variances of activities that are common to both paths.

The transpose of the vector, $\underline{P}'$, of the paths completion times:

$$\underline{P}' = [P_1, \dots, P_n]$$

will have a multivariate normal distribution with a symmetric variance-covariance matrix $\underline{A}$, where

$$\underline{A} = [\sigma_{ij}^2],$$

and with a vector of mean $\underline{U}'$, where

$$\underline{U}' = [\mu_1, \dots, \mu_n].$$

The distribution function, $f(\underline{P}, \underline{U}, \underline{A}, n)$, appears extensively in the statistics literature and is as follows:

$$f(\underline{P}, \underline{U}, \underline{A}, n) = \frac{\exp\{-\frac{1}{2}(\underline{P} - \underline{U}')' \underline{A}^{-1} (\underline{P} - \underline{U})\}}{(2\pi)^{\frac{n}{2}} |\underline{A}|^{\frac{1}{2}}}. \tag{1}$$

It is interesting to note that equation (1) contains all the information required to solve the PERT network completion time problem under the normal assumption. The analytical solution requires a transformation of equation (1) that will yield the mean and variance, or better still the distribution, of the maximum element of the vector $\underline{P}$. [6] was able to find the exact mean and variance in the case when $n = 2$, but the general solution, $n > 2$, appears to have escaped researchers so far.

AN APPROXIMATE SOLUTION

Let N_i, $i = 1, \dots, n$, be normal random variables with means μ_i and standard deviations σ_1^2, let E_n and V_n, respectively, be the mean and variance of maximum $(N_1, \dots, N_n)$, and let $r(X, Y)$ be the coefficient of linear correlation between any two normal variables, X and Y. The density and cumulative density functions, $\psi(\alpha)$ and $\phi(\alpha)$, of a standard normal variable are as follows:

$$\psi(\alpha) = \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{\alpha^2}{2}\right) \quad \text{and} \quad \phi(\alpha) = \int_{-\infty}^{\alpha} \psi(t)dt.$$

[6] proved the following expressions for E_2 and V_2:

$$E_2 = \mu_1\, \phi(\alpha_2) + \mu_2\, \phi(-\alpha_2) + a_2\, \psi(\alpha_2) \tag{2}$$

$$V_2 = (\mu_1^2 + \sigma_1^2)\, \phi(\alpha_2) + (\mu_2^2 + \sigma_2^2)\, \phi(-\alpha_2) + (\mu_1 + \mu_2) a_2\, \psi(\alpha_2) - E_2^2, \tag{3}$$

where

$$a_2^2 = \sigma_1^2 + \sigma_2^2 - 2\,\sigma_1\,\sigma_2 r(N_1, N_2) \quad \text{and} \quad \alpha_2 = \frac{\mu_1 - \mu_2}{a_2}.$$

He also proved that

$$r(\max(N_1, N_2), N_3) = \frac{\sigma_1\, r(N_1, N_3)\, \phi(\alpha_2) + \sigma_2\, r(N_2, N_3)\, \phi(-\alpha_2)}{\sqrt{V_2}}. \tag{4}$$

Good approximations for E_n and V_n were obtained by [6] by assuming, the $\max(N_1, \dots, N_{n-1})$ to be a normal random variable. With this assumption, expressions (2) and (3) can be turned into recursive expressions, thus enabling the computation of the mean and variance of $\max[\max(N_1, \dots, N_{n-1}) N_n]$. The following substitutions are made:

$$\begin{array}{ll} E_2 \to E_n, & V_2 \to V_n, \\ \mu_1 \to E_{n-1}, & \sigma_1^2 \to V_{n-1}, \\ \mu_2 \to \mu_n, & \sigma_2^2 \to \sigma_n^2, \end{array}$$

to arrive at the following general expressions for $n \geq 2$:

$$E_n = E_{n-1}\, \phi(\alpha_n) + \mu_n\, \phi(-\alpha_2) + a_n\, \psi(\alpha_n), \tag{5}$$

$$V_n = (E_{n-1}^2 + V_{n-1})\, \phi(\alpha_n) + (\mu_n^2 + \sigma_n^2)\, \phi(-\alpha_n) + (E_{n-1} + \mu_n)\, a_n\, \psi(\alpha_n) - E_n^2, \tag{6}$$

where

$$a_n^2 = V_{n-1} + \sigma_n^2 - 2\sqrt{V_{n-1}}\, \sigma_n\, r[\max(N_1, \dots, N_{n-1}), N_n]$$

and

$$\alpha_n = \frac{E_{n-1} - \mu_n}{a_n}.$$

The term $r[\max(N_1, \dots, N_{n-1})]$ still needs to be computed to generalise the results. The expression $r[\max(N_1, \dots, N_{n-1})]$ is, in fact, the same as $r(\max[\max(N_1, \dots, N_{n-2}), N_{n-1}], N_n)$. The assumption that the maximum of two normal random variables is also normal will allow equation (4) to be transformed into a recurrence relation, as follows:

$$r[\max(N_1, \dots, N_k), N_n] = \frac{\sqrt{V_{k-1}}\, r[\max(N_1, \dots, N_{k-1}), N_n]\, \phi(\alpha_k) + \sigma_k r(N_k, N_n) \phi(-\alpha_k)}{\sqrt{V_k}}. \tag{7}$$

The dummy variable k can run from 2 to $n - 1$, $(n > 3)$, to obtain $r[\max(N_1, \dots, N_{n-1}), N_n]$.

The recurrence relationships, (5)-(7) can be used to compute an approximation to the mean and standard deviation of the maximum of a set of correlated normal variables.

EXAMPLES WITH NORMAL ACTIVITIES

Two example networks are considered in which the durations of all activities are assumed to be normally distributed. Expressions (5)-(7) are initialized by setting $E_1 = \mu_1$ and $V_1 = \sigma_1^2$ and are applied in a recursive manner to find E_n and V_n; i.e., the n paths are introduced one by one until all n paths have been considered. The path with the longest mean duration time is introduced first, then the second longest, etc. However, there are many ways of selecting paths. Dealing with paths that have the long mean durations first will allow a cut-off in terms of the accuracy required after a few paths have been considered, thus eliminating the need to consider the many hundreds of paths that are often found in "real life" networks. The paths with the longest mean durations are, other factors being equal, more likely to become the longest path of the network when an actual trial is realized.

The first example network is shown in Figure 1. It consists of nine paths, and the mean duration times of each path are progressively smaller (see Table 1 for results). The second example is taken from [13], with the mean duration time of an activity equal to three times its standard deviation. This network has five paths all having the same mean duration time (see Table 2) and is likely to cause the most serious problems, in terms of the main assumption, that the maximum of a number of normal variables is a normal variable. The results of these two examples are shown in Tables 3 and 4, respectively. The mean and standard deviations of the completion times of these

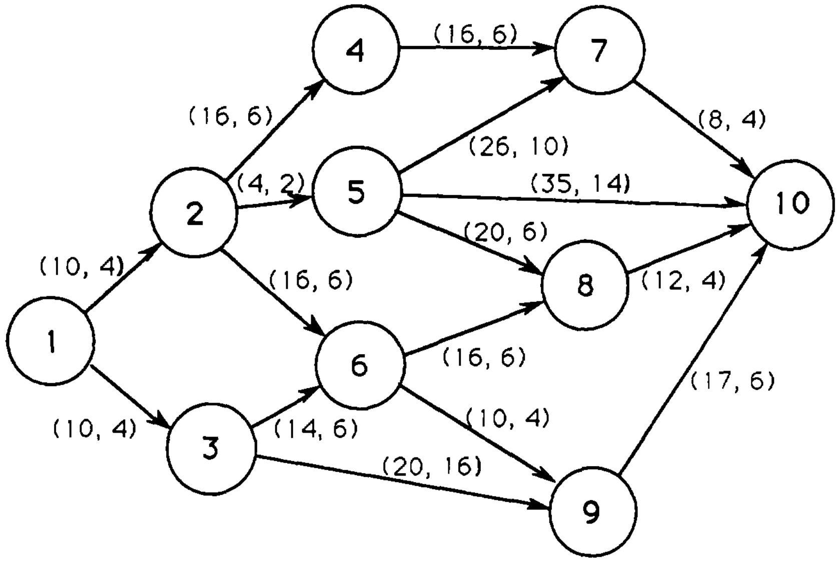

Figure 1. Example network. Figures in brackets are the mean and variance of the respective activity duration time.

networks were also obtained via simulation. The simulation results can be used to ascertain the accuracy of the proposed approximation. It can be seen that the approximation is very close when compared to the simulation results (see Table 5). This is true for both the mean and standard deviations of the completion times of the example networks.

Table 1. Vector of means and variance-covariance matrix for the network in Figure 1.

Paths	Vector of means, $\underline{U}$	Variance-covariance matrix $\underline{A}$ P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	Path nodes
P_1	54	20	10	10	0	4	4	4	0	8	1 - 2 - 6 - 8 - 10
P_2	53	10	20	0	10	4	4	4	6	4	1 - 2 - 6 - 9 - 10
P_3	52	10	0	20	10	0	0	0	4	4	1 - 3 - 6 - 8 - 10
P_4	51	0	10	10	20	0	0	0	10	0	1 - 3 - 6 - 9 - 10
P_5	50	4	4	0	0	20	4	8	0	4	1 - 2 - 4 - 7 - 10
P_6	49	4	4	0	0	4	20	6	0	6	1 - 2 - 5 - 10
P_7	48	4	4	0	0	8	6	20	0	6	1 - 2 - 5 - 7 - 10
P_8	47	0	6	4	10	0	0	0	26	0	1 - 3 - 9 - 10
P_9	46	8	4	4	0	4	6	6	0	16	1 - 2 - 5 - 8 - 10

Table 2. Vector of means and variance-covariance matrix for the network in [13].

Paths	Vector of means, $\underline{U}$	Variance-covariance matrix $\underline{A}$ P_1	P_2	P_3	P_4	P_5	Path nodes
P_1	12	6	4	0	0	0	1 - 2 - 5 - 8
P_2	12	4	6	1	0	0	1 - 3 - 5 - 8
P_3	12	0	1	6	1	0	1 - 3 - 6 - 8
P_4	12	0	0	1	6	4	1 - 4 - 6 - 8
P_5	12	0	0	0	4	6	1 - 4 - 7 - 8

Table 3. Results for example network shown in Figure 1.

Number of paths included	Mean of newly introduced path	Variance of newly introduced path	Mean of completion time: Simulation N = 10000	Mean of completion time: Analytical	Mean of completion time: Percentage error to simulation	Standard deviation of completion time: Simulation N = 10000	Standard deviation of completion time: Analytical	Standard deviation of completion time: Percentage error to simulation
P_1	54	20	54.04	54.00	0.07	4.50	4.47	0.67
P_1 to P_2	53	20	55.36	55.33	0.05	4.12	4.11	0.24
P_1 to P_3	52	20	56.07	56.09	0.04	3.81	3.82	0.26
P_1 to P_4	51	20	56.31	56.45	0.25	3.75	3.66	0.40
P_1 to P_5	50	20	56.61	56.74	0.23	3.56	3.49	1.97
P_1 to P_6	49	20	56.77	56.90	0.23	3.46	3.39	2.02
P_1 to P_7	48	20	56.83	56.97	0.25	3.42	3.34	2.34
P_1 to P_8	47	26	56.91	57.05	0.25	3.39	3.30	2.65
P_1 to P_9	46	16	56.91	57.06	0.26	3.38	3.29	2.66

Table 4. Results for example network shown in [13].

Number of paths included	Mean of newly introduced path	Variance of newly introduced path	Mean of completion time: Simulation N = 10000	Mean of completion time: Analytical	Mean of completion time: Percentage error to simulation	Standard deviation of completion time: Simulation N = 10000	Standard deviation of completion time: Analytical	Standard deviation of completion time: Percentage error to simulation
P_1	12	6	11.97	12.00	0.25	2.46	2.45	.041
P_1 to P_2	12	6	12.81	12.80	0.08	2.32	2.32	0
P_1 to P_3	12	6	13.72	13.72	0	2.01	2.01	0
P_1 to P_4	12	6	14.24	14.26	0.14	1.86	1.84	1.08
P_1 to P_5	12	6	14.52	14.56	0.28	1.81	1.77	2.21

Table 5. Comparison of different methods.

Method used	Network in Figure 1: Mean	Network in Figure 1: Standard deviation	Network in [13]: Mean	Network in [13]: Standard deviation
Simulation N = 10000	56.91	3.38	14.52	1.81
Analytical method presented in this paper	57.06	3.29	14.56	1.77
Percentage error to simulation	0.26	2.66	0.28	2.21
Sculli [8]	57.44	2.95	14.69	1.68
Percentage error to simulation	0.93	12.7	1.17	7.18
PERT-calculated	54.00	4.47	12.00	2.45
Percentage error to simulation	5.11	32.2	17.4	35.4

APPROXIMATIONS TO NON-NORMAL ACTIVITIES

In a general sense, the proposed approximation is simply a series of formulas connecting the mean and standard deviations of a number of statistical distributions. The approximation can also be applied when the distributions of individual activities are not normal, the only requirement being that the means and standard deviations are known. Two examples are used to explore the usefulness of the approximation when nonnormal individual activity durations are involved. Both examples use the same network, taken from [13].

The first example assumes all activities to follow the same negative exponential distribution:

$$f(x) = \exp(-x/3)/3 \qquad 0 < x < \infty,$$

$$\mu = 3, \quad \sigma = 3.$$

The second example assumes all activities to follow the same uniform distribution:

$$f(x) = 1/12 \qquad 0 < x < 12,$$

$$\mu = 6, \quad \sigma = 2/\sqrt{3}.$$

The results are shown in Table 6. It can be seen that the error in the mean is insignificant in both examples, despite the large asymmetry of the negative exponential distribution. The error in the standard deviation for the negative exponential case is unacceptable, while that for the uniform distribution case is acceptable. The unacceptable error in the case of the negative exponential is most likely due to the wide dispersion of this distribution from its mean. However, all paths in these two sample networks contain three activities only. This number is too small for the Central Limit Theorem to have any significant impact on the durations of paths. As the number of activities on a path increases, the error due to the specific nature of the distributions of individual activities will diminish.

Table 6. Error caused in approximation to normal function.

Network completion time	Negative exponential		Uniform	
	Mean	Standard deviation	Mean	Standard deviation
Analytical method presented in this paper	14.61	3.68	24.48	4.25
Simulation N = 10000	14.83	5.32	24.50	4.14
Percentage error to simulation	1.48	30.8	0.08	2.66

CONCLUDING COMMENTS

The results of this paper are encouraging, particularly since the computational and programming efforts required to simulate a network are very much greater than the proposed approximation. It is also true that simulation of an entire network is not practical, or even possible, when the network is large (see [10]).

It also seems that it is generally correct to conclude that the more symmetric the distributions of activities, the smaller the error. When the longer paths of a network contain several activities, the Central Limit Theorem will push the duration time of the path toward normality and symmetry. Networks in "real life" often contain several hundred or even thousands of activities. The duration of a path connecting the first and last event of "real life" networks will depend on the sum of the many activities along it, and will tend toward normality irrespective of the distributions of individual activities along it.

On a concluding note, the majority of research on estimation of the completion time of PERT networks assumes that some information is known, such as the mean and variance, or the exact distribution, of the duration time of individual activities, and indeed, this must be so. However, there is the perennial discussion as to why projects are always late; see [14,15]. The problem stems from what appears to be an inability to accurately estimate the completion time distribution of individual activities. A considerable number of fundamental questions are being reconsidered [11,16], particularly on the use of the Beta distribution and its three time estimates, as suggested by [17]. This indicates that there is still considerable research potential in both the areas of estimating individual activity duration times and the overall time analysis of networks.

REFERENCES

1. P. Robillard and M. Trahan, The completion time of PERT networks, *Ops. Res.* **25**, 15–29 (1977).
2. G.B. Kleindorfer, Bounding distributions for stochastic acyclic networks, *Ops. Res.* **19**, 1586–1601 (1971).
3. G.B. Kleindorfer and P.R. Kleindorfer, Bounding distributions for stochastic logic, *Opl. Res.* Q.**25**, 465–479 (1974).
4. D.R. Fulkerson, Expected critical path lengths in PERT networks, *Ops. Res.* **10**, 808–817 (1962).
5. S.E. Elmaghraby, On the expected durations of PERT networks, *Mgmt. Sci.* **13**, 299–306 (1967).
6. C.E. Clark, The greatest of a finite set of random variables, *Ops. Res.* **9**, 145–162 (1961).
7. W.R. Greer and La Cava, Approximation for the greater of two normal variables, *Omega* **7**, 361–363 (1979).
8. D. Sculli, The completion time of PERT networks, *J. Opl. Res.* **34**, 155–158 (1983).
9. R.M. Van Glyke, Monte Carlo methods and the PERT problem, *Ops. Res.* **11**, 839–860 (1963).
10. T.M. Cook and R.H. Jennings, Estimating a project's completion time distribution using intelligent simulation methods, *J. Opl. Res.* **30**, 1103–1109 (1979).
11. D. Sculli, A historical note on PERT times, *Omega* **17**, 195–196 (1989).
12. P. Kaiomars, Asnklesaria, and Z.V.I. Dregner, A multivariate approach to estimating the completion time of PERT networks, *J. Opl. Res.* **37**, 811–815 (1986).
13. J. Kamburoski, Normal distributed activity durations in PERT networks, *J. Opl. Res.* **36**, 1051–1057 (1985).
14. R.J. Schonberger, Why projects are "always" late: A rational based on manual simulation of PERT/CPM networks, *Interfaces* **11**, 66–70 (1981).
15. D. Sculli and K.L. Wong, The maximum and sum of two Beta variables and the analysis of PERT networks, *Omega* **13**, 233–240 (1985).
16. M.W. Sasieni, A note on PERT times, *Mgmt. Sci.* **32**, 1652–1653 (1986).
17. C.E. Clark, The PERT model for the distribution of an activity time, *Ops. Res.* **10**, 405–406 (1962).